Intron mobility in phage T4 is dependent upon a distinctive class of endonucleases and independent of DNA sequences encoding the intron core: mechanistic and evolutionary implications

Deborah Bell-Pedersen[1,2], Susan Quirk[1], Jonathan Clyman[1] and Marlene Belfort[1,3]
[1]Wadsworth Center for Laboratories and Research, New York State Department of Health, Empire State Plaza, PO Box 509, Albany, NY 12201-0509, [2]Biology Department, State University of New York at Albany, Albany, NY 12222 and [3]School of Public Health, State University of New York at Albany, Empire State Plaza, Albany, NY 12201, USA

Received April 12, 1990; Accepted May 10, 1990

ABSTRACT

Although mobility of the phylogenetically widespread group I introns appears to be mechanistically similar, the phage T4 intron-encoded endonucleases that promote mobility of the *td* and *sun*Y introns are different from their eukaryotic counterparts. Most notably, they cleave at a distance from the intron insertion sites. The *td* enzyme was shown to cleave 23 – 26 nt 5′ and the *sun*Y endonuclease 13 – 15 nt 3′ to the intron insertion site to generate 3-nt or 2-nt 3′-OH extensions, respectively. The absolute coconversion of exon markers between the distant cleavage and insertion sites is consistent with the double-strand-break repair model for intron mobility. As a further critical test of the model we have demonstrated that the mobility event is independent of DNA sequences that encode the catalytic intron core structure. Thus, in derivatives in which the *lac*Z or *kan*^{R} coding sequences replace the intron, these marker genes are efficiently inserted into intron-minus alleles when the cognate endonuclease is provided *in trans*. The process is therefore endonuclease-dependent, rather than dependent on the intron *per se*. These findings, which imply that the endonucleases rather than the introns themselves were the primordial mobile elements, are incorporated into a model for the evolution of mobile introns.

INTRODUCTION

The group I introns constitute a phylogenetically widespread class of intervening sequences characterized by both structural and functional uniformity (reviewed in refs. 1 and 2). These introns have in common a secondary structure that forms the catalytic core required for splicing (2–5). In many cases the core structures are interrupted by open reading frames (ORFs) (1). In one subclass of group I introns the ORFs encode endonucleases, which endow the introns with mobile properties (reviewed in refs. 6, 7). Specifically, the endonucleases catalyse a double-strand break in intronless (In^-) cognates of the intron-containing (In^+) alleles. The cleavage has been proposed to stimulate recombination via a double-strand-break repair (DSBR) mechanism (8, 9), which was initially advanced as a model for gene conversion in yeast (10, 11). According to the model, the DNA ends undergo exonucleolytic degradation prior to invading the homologous In^+ donor strands, initiating a repair process that results in efficient conversion of the In^- recipient to In^+ (Fig. 1). This process is termed 'homing', reflecting intron acquisition by an In^- variant of the same or closely related gene (12). Although the model satisfactorily accounts for observed features of the intron homing process, including coconversion of flanking exon sequences, (9, 13–16), it remains to be experimentally verified.

The ω intron in the large rRNA gene of *Saccharomyces cerevisiae* mitochondria (17) provided the first example of intron homing. The dependence of the process on an ORF-encoded endonuclease was subsequently established (9, 13, 18, 19). There are now five additional examples of endonuclease-encoding mobile introns in widely diverse organisms and cellular compartments. These include the aI4α intron in the *cox*1 gene of *S. cerevisiae* (15, 20), the LSU-5 intron in *Chlamydomonas eugametos* chloroplasts (21 and C. Lemieux, personal communication), a nuclear intron in the *Physarum polycephalum* large rRNA gene (16) and the *td* and *sun*Y introns of phage T4 (14, 22, 23).

Although the overall mobility process seems similar for the pro- and eukaryotic introns (6, 7), the endonucleases encoded by the T4 *td* and *sun*Y introns (I-TevI and I-TevII, respectively; for nomenclature see ref. 12) have different properties from their eukaryotic counterparts in the *S. cerevisiae* (ω, aI4α) and *P. polycephalum* rRNA introns (15, 20, 24–26). First, I-TevI and I-TevII appear to have a more relaxed site specificity than the eukaryotic enzymes (22 and D. B-P., unpublished), which have recognition sequences as long as 18 nt (6, 7, 24). Second, unlike the yeast endonucleases, the phage enzymes do not contain the conserved dodecapeptide (LAGLI-DADG) motif (3, 27), indicating that they may represent a different family of

endonucleases. Third, the results reported here reveal additional distinctive properties of the T4 enzymes, the most striking of which is that their cleavage sites are distant from the intron insertion sites. In contrast, the ω, aI4α and *P. polycephalum* endonucleases all cleave at or within one nucleotide of the intron insertion site.

Considered in the context of the DSBR model for intron homing, the eccentric cleavages of the phage endonucleases lead to the prediction that exon sequences between the cleavage site and the intron must be coinherited (Fig. 1). Accordingly, we have observed 100% coconversion of markers that lie between the *td* cleavage and insertion sites (14). A further prediction of the model is that the DNA encoding the intron core is a passive element in the homing process, instead of contributing specific information, as is the case for other types of mobile elements. In the present study, we demonstrate that the recombination events are indeed independent of catalytic RNA sequences of the intron, again supporting the DSBR model, while providing evidence that the endonucleases rather than the introns themselves may have been the original mobile elements. These findings are incorporated into a model for the evolution of mobile introns.

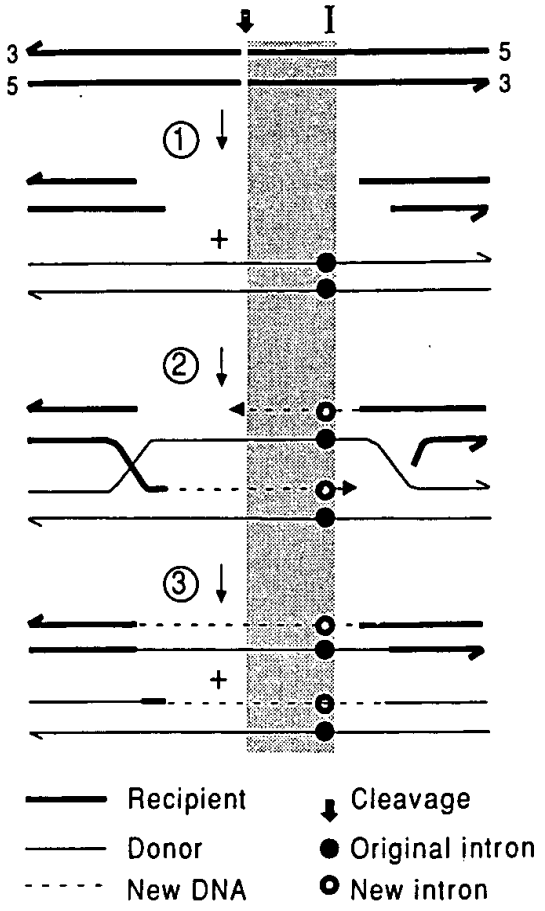


Figure 1. The Double-strand-break repair (DSBR) model for intron homing. The DSBR model (11) is shown with adaptations to *td* intron inheritance. Initially, a double-strand cleavage (heavy arrow) occurs upstream of the intron insertion site (I) in the phage recipient. This break is enlarged to a gap, with exonucleolytic degradation extending past the intron insertion site (step 1). After alignment of donor and recipient, the 3′ ends invade the intron-containing donor molecule, initiating D-loop formation. Both 3′ ends act as primers for donor-directed repair synthesis (step 2). Following ligation, and resolution of the double Holliday crossovers (only the non-recombinant mode is shown), each duplex harbors an identical copy of the intervening sequence (step 3). The shading designates obligatory coconversion of exon sequences between the cleavage site and the intron.

MATERIALS AND METHODS

Media

Bacteria were grown in TB (1% Bacto-tryptone [Difco] and 0.5% NaCl) or in TBYE (TB containing 0.5% yeast extract). Antibiotics were added where appropriate: ampicillin and kanamycin at 100 μg/ml and tetracycline at 10 μg/ml. Bacterial lawns were plated in a top agar overlay in TB containing 0.7% Bacto-agar on TB plates (TB containing 1.1% Bacto-agar).

Phages

T4*td*ΔIn contains a precise deletion of the *td* intron (14). T4SAΔ9*td*ΔIn, the phage used as recipient for the *lac*Z-marked intron (Fig. 4), is a derivative of T4SAΔ9 (obtained from D. Shub). The 2.4-kb SAΔ9 deletion enables the phage to package its DNA following insertion of the 4-kb *lac*Z fragment. *td*ΔIn is λgt11*td*ΔIn, which was constructed by cloning the 1.85 kb *Eco*RI *td*ΔIn fragment (28) into the *Eco*RI arms of λgt11 (Promega). M13 constructs were made by inserting the *Eco*RI *td*ΔIn fragment into the *Eco*RI site of M13mp11 to generate both the sense (M13*td*ΔIn) and antisense (M13*td*ΔIn-as) single strands. M13*sun*YΔIn constructs were made by cloning the *Pst*I-*Xba*I *sun*YΔIn fragment (29) into either M13mp18 (for the antisense strand) or M13mp19 (for the sense strand).

Bacterial Strains and Plasmids

Plasmid pA*td*SG2 is pACYC184 (30) with the *td* gene (*Eco*RI fragment) containing a truncated 265-nt intron (ΔP6−2, ref. 31). This construct has an intact core structure, but is deleted for ORF sequences, which are replaced by a *Sma*I site to facilitate cloning. Plasmids pA*td*SG2*lac*Z and pA*td*SG2*kan*^{R} contain a 4-kb *Sma*I-*Dra*I *lac*Z fragment derived from pNK683 (32) or the 1.2-kb *Hin*cII *kan*^{R} fragment from pUC4K (Pharmacia) in the *Sma*I site, respectively. For construction of pA*td*ΔIn*lac*Z and pA*td*ΔIn*kan*^{R}, a unique *Sma*I site was engineered at the site of intron insertion into *td*ΔIn by oligonucleotide-directed mutagenesis (Table 1, oligomer 5). M13mp11*td*ΔIn single-stranded DNA synthesized in RZ1032 (*dut*1, *ung*1, *sup*E44; ref. 33) was used to enrich for the mutagenized strand following transformation into JM101 (*sup*E, *thi*, [F′ *tra*D36, *pro*AB, *lac*I^{q}ZΔM15]). The resulting *td*ΔIn-*Sma*I fragment was cloned into the *Eco*RI site of pACYC184 and used for cloning the *lac*Z and *kan*^{R} fragments as above, to generate pA*td*ΔIn*lac*Z and pA*td*ΔIn*kan*^{R}, respectively. The ORF donor plasmids pK-ORF$^+$ and pK-ORF$^-$ are pKC30 derivatives (34) containing an intact ORF (the *td*P-P fragment, ref. 22) or a disrupted ORF (the *td*Δ1−3 fragment, ref. 22), respectively.

Genetic Analyses

Intron mobility assays were performed as described in ref. 22 for plasmid to T-even phage crosses (Fig. 4A). N99(λcI857*bio*10*P*am3) was transformed with the Amp^R pK-ORF$^+$ or pK-ORF$^-$ plasmids. These lysogens were cotransformed with compatible Tet^R plasmids containing either the *td* intron core plus *lac*Z (pA*td*SG2*lac*Z) or *lac*Z inserted between the two exons (pA*td*ΔIn*lac*Z). Cotransformants were infected with T4SAΔ9*td*ΔIn by the plate method (22), incubated

at 37°C for 2 h to induce ORF expression from the λP_L promoter on pKC30, and then incubated overnight at 30°C. Lac$^+$ progeny phage were detected by plating on X-gal indicator plates.

Plasmid to λ*td*ΔIn crosses were performed in Y1089 (Δ*lac*U169, *pro*A$^+$, Δ*lon*, *str*A, *hfl*A [chr::tn10] r_k^-,m_k^-, ref. 35) made lysogenic for λ*td*ΔIn (Fig. 4B). The lysogens were used as hosts for pK-ORF$^+$ or pK-ORF$^-$ plasmids and were cotransformed with pA*td*SG2*kan*^{R} or pA*td*ΔIn*kan*^{R}. After growth to mid-log phase at 30°C in TBYE plus ampicillin and tetracycline, the lysogens were induced by shifting cultures to 44°C for 15 min, followed by growth at 38°C for 4 h. $CHCl_3$ was added to lyse the bacterial cells. Following amplification, λ progeny phage were used to make lysogens of Y1089. The frequency of Kan^R lysogens was calculated from the number of Kan^R lysogens relative to the total number of lysogens plated (as scored by surviving colonies on a plate seeded with λ*c*I$^-$ phage).

Endonuclease Preparations

In vitro translation of the *td* endonuclease, I-TevI, was as described (14) except that for this work *td* transcripts were translated in wheat-germ extracts (Promega), rather than in reticulocyte lysates. The *sun*Y endonuclease, I-TevII, was synthesized in rabbit reticulocyte extracts (Promega). Transcription was performed on 2 μg of *Xba*I-linearized pSP64-*sun*Y (a *Pst*I-*Xba*I *sun*Y fragment [29] cloned into vector pSP64). The SP6 polymerase-generated RNA templates were extracted once with phenol, ethanol precipitated and translated *in vitro* according to the manufacturer's instructions.

Cleavage Site Mapping

Mapping of endonuclease cleavage sites was essentially by the method of Wenzlau et al (15). M13mp11*td*ΔIn, M13mpll*td*ΔIn-as, M13mp18*sun*YΔIn or M13mp19*sun*YΔIn DNA (2μg) were annealed to 20 pmol of sequencing primer in a 20 μl sequencing reaction (Sequenase, US Biochemicals). One-fifth of this reaction was added to 2.5 μl of 1 mM dNTPs to generate the double-stranded substrate DNA. The remainder of the reaction was split and incubated with ddNTPs to generate the sequencing ladders. The double-stranded substrate was incubated with 5 μl of translation extract (+/− endonuclease) at 25°C, in the presence of 10 mM $MgCl_2$, 50 mM Tris-HCl pH 8.0, 100 mM NaCl. The cleavage reactions were stopped by the addition of SDS to 1%. Single-stranded salmon sperm DNA (0.5 μg) was added prior to two successive phenol:chloroform/isoamyl alcohol extractions. Following ethanol precipitation, the samples were heated for 15 min at 70°C, and then separated on a 6% acrylamide-7 M urea sequencing gel.

RESULTS

I-TevI and I-TevII Cleavage Sites are Remote from the Intron Insertion Sites

The cleavage sites of I-TevI and I-TevII were mapped on the respective *td* and *sun*Y In$^-$ target DNAs as described by Wenzlau et al (15) and shown to be distant from the intron insertion sites. Single-stranded In$^-$ DNA was primed downstream of the insertion site on the top strand and upstream of the insertion site on the complementary strand to generate double-stranded DNA cleavage substrates. The precise cut sites were determined by separating the cleavage products alongside

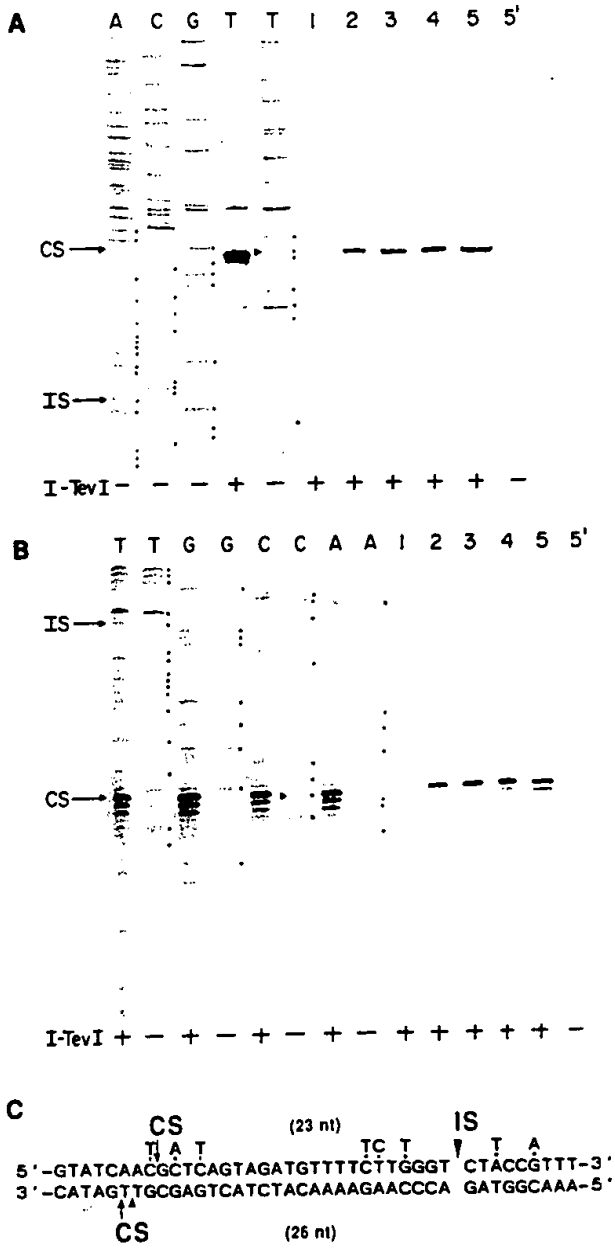


Figure 2. The *td* intron endonuclease (I-TevI) cleavage sites. (A) Bottom-strand cut site. M13*td*ΔIn single-stranded DNA, primed downstream of the intron insertion site (oligomer 1, Table 1), was extended with Sequenase (US Biochemicals) in the presence of dNTPs (including ^{35}S-dATP). The resulting double-stranded DNA substrate was cleaved with I-TevI, synthesized in wheat germ extracts, for 1.0, 1.5, 2.0, 2.5 and 5.0 min (lanes 1–5, respectively). Lane 5' shows substrate incubated with unprogrammed wheat germ extract. Sequencing ladders were generated using the corresponding chain-terminating dideoxynucleotide. The cleavage site (CS) was precisely determined by adding cleaved substrate (5 min digestion) to the sequence ladder prior to gel electrophoresis (first T lane). The CS and the intron insertion site (IS) are indicated with arrows. (B) Top-strand cut site. M13*td*ΔIn-as single-stranded DNA, primed upstream of the intron insertion site (oligomer 2, Table 1) was treated as in Panel A. Cleaved DNA from the 5-min timepoint was coelectrophoresed in the first of each sequencing lane. (C) The T4*td*ΔIn cleavages generated by I-TevI. For the top strand, only the primary cleavage seen in early timepoints is indicated. The arrowhead at −25 identifies a primary cut site observed in some experiments utilizing double-stranded plasmid DNA templates (data not shown). The distance of the primary CS from the IS for both strands is indicated in parentheses. Sequence variations in In$^-$ phage T2L (36) are indicated above dots, whereas single-site mutations in T4*td*ΔIn (14) are indicated above dashes. All these sequence polymorphisms are consistent with efficient intron inheritance.

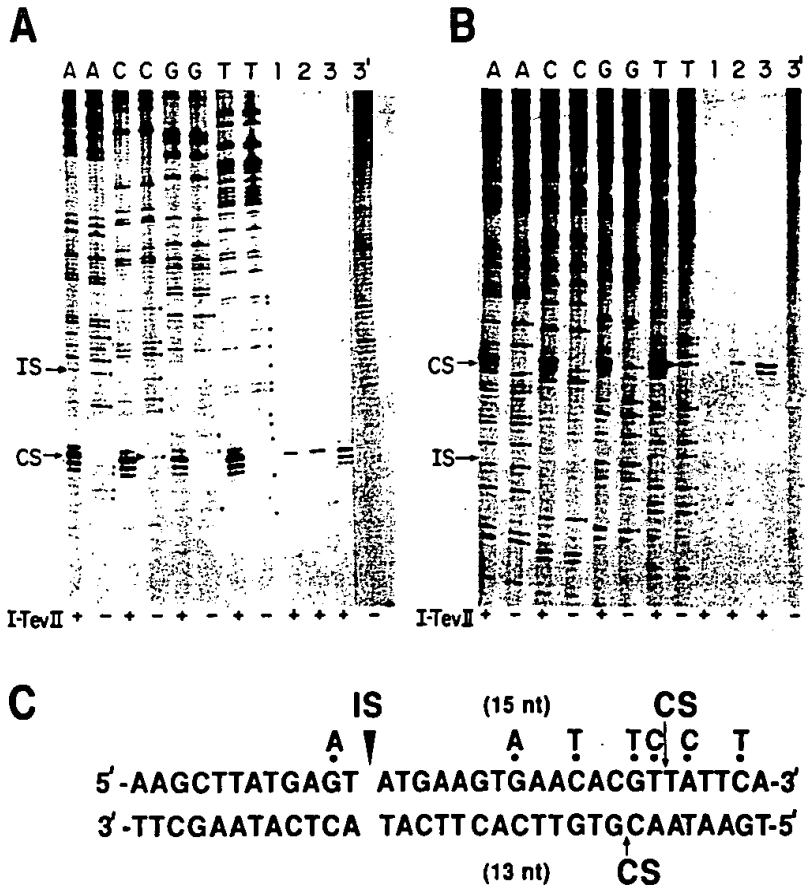


Figure 3. The *sun*Y intron endonuclease (I-TevII) cleavage sites. Cleavage and sequencing reactions were performed as in Fig. 2, where panels A and B show M13mp19*sun*YΔIn DNA primed downstream of the intron insertion site (oligomer 3, Table 1), and M13mp18*sun*YΔIn primed upstream of the insertion site (oligomer 4, Table 1), respectively. Lanes 1–3 show the double-stranded substrate cleaved for 1, 2 and 5 min, respectively, with I-TevII, which was synthesized in rabbit reticulocyte-lysates (Promega). Lane 3' shows the same substrate DNA incubated for 20 min in unprogrammed rabbit reticulocyte lysate. Cleaved DNA from the 5 minute timepoint was coelectrophoresed in the first of each sequencing lane. Both the primary cleavage site (CS), and intron insertion site (IS) are indicated with arrows to the left of the autoradiograms and above the sequence in panel C. Labeling in panel C is as in Fig. 2, with sequence polymorphisms in T2L indicated above dots.

Table 1: Oligonucleotides used in this study.

#	Description	Nt number	Sequence
1	*td* exon II	1879–1890[1]	5'-GATATGAGTATTACCACCA-3'
2	*td* exon I	671–689[1]	5'-TGGCATTACCGCCTTGTCA-3'
3	*sun*Y exon II	3960–3980[2]	5'-CGAACATAACAAGCTTAGGGA-3'
4	*sun*Y exon I	2701–2718[2]	5'-TCGCTGATGCGTTGAATT-3'
5	*td* mutagenic-*Sma*I[3]		5'-ATTAAACGGTAGGATCCCGGGTACCCAAGAAAAC-3'

[1] Numbering as per ref. 46
[2] Numbering as per ref. 39
[3] The inserted nucleotides containing the *Sma*I site are underscored

(or together with) a sequencing ladder produced from the same DNA (Figs. 2 and 3).

For *td* I-TevI synthesized *in vitro*, cleavage was observed 23 nt upstream of the intron insertion site on the sense strand, and 26 nt 5' to the intron insertion site on the antisense strand, generating a 3 nt 3' extension (Fig. 2). While the primary cleavage site was verified through numerous repetitions of these experiments, we sometimes observed an additional antisense cleavage site as indicated by an arrowhead on Fig. 2C. These results may reflect reduced specificity of cleavage on the bottom strand. The appearance of shorter cleavage products with increasing times of incubation (Fig. 2B) may also result from relaxed cleavage specificity, although exonucleolytic degradation has not been ruled out. Regardless, identical primary cleavages were observed with I-TevI that had been overexpressed in *Escherichia coli* and partially purified (data not shown), verifying the faithfulness of the *in vitro*-synthesized protein. Additionally, cleavage generates 3'-OH and 5'-PO_4 ends, as evidenced by the ability of cleaved DNA to be ligated by T4 DNA ligase (data not shown).

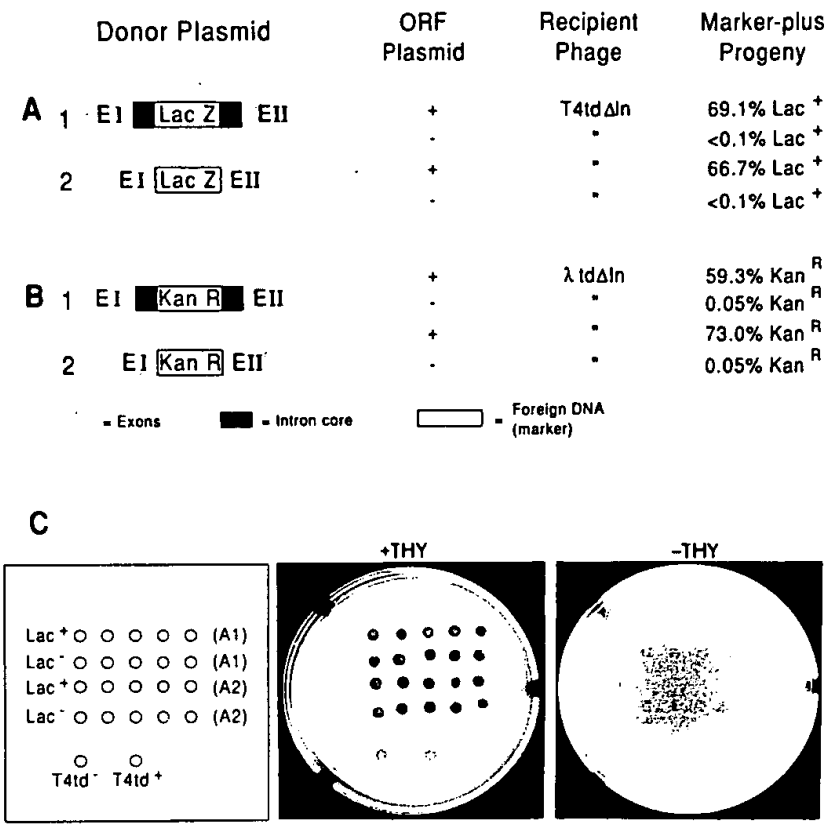


Figure 4. Intron core-independent gene conversion. Panels A and B show *lac*Z and *kan*^{R} inheritance frequencies, respectively, in the presence (constructs 1) or absence (constructs 2) of intron core sequences. The presence (+) or absence (−) of an intact ORF on a compatible plasmid is indicated. The frequency of *lac*Z inheritance by T4SAΔ9*td*ΔIn was determined by quantitating the ratio of blue plaques to total progeny phage on X-gal indicator plates. The frequency of *kan*^{R} inheritance by λ*td*ΔIn was established by quantitating the ratio of Kan^R lysogens to total lysogens. Phenotypic analysis of progeny from the crosses in panel A is shown in panel C. Lac$^-$ and Lac$^+$ phage were tested for the Td phenotype by their ability to produce plaques on *thy*A$^-$GM201 in the presence (+THY) or absence (-THY) of thymine (47). Progeny phage isolated from cross A1 appear in the top two rows, whereas progeny phage from cross A2 appear on the following two rows, as per the template on the left. T4*td*$^-$ and T4*td*$^+$ control phages are shown below.

The *sun*Y I-TevII cleavage site is also displaced from the insertion site (Fig. 3), but in this case is located 3′ to the intron insertion site. Cleavage was observed 15 nt downstream of the intron insertion site on the sense strand, and 13 nt 3′ of the insertion site on the antisense strand, generating a 2-nt 3′ extension. These results emphasize the distinctive nature of the phage endonucleases: First, the cleavages are distant from the intron insertion sites in either the 5′ or 3′ direction, rather than in close proximity. Second, the 2- and 3-nt 3′ extensions are shorter than those of the *S. cerevisiae* and *P. polycephalum* endonucleases, which are 4 nt in length (15, 20, 24, 26). Third, the phage enzymes are tolerant of extensive sequence heterogeneity around the intron insertion and endonuclease cleavage sites (Figs. 2C and 3C). The sequence polymorphisms, which occur in the *td* and *sun*Y homing sites of the related T2 and T4 phages (22, 36) and in specific T4*td* mutants (14), are consistent with function of their cognate enzymes, I-TevI and I-TevII, contrasting with the extreme site-specificity of the eukaryotic enzymes (6, 7, 24).

DNA Sequences Encoding the Intron Core do not Contribute Information Required for Mobility

Despite the distinctive features of the phage-encoded endonucleases, the mechanism of intron transfer appears similar in pro- and eukaryotes (reviewed in refs. 6, 7, 37). The unidirectionality of intron transfer, as well as the coconversion of flanking exons, a reflection of exonuleolytic degradation, are both common features that are readily accommodated by the double-strand-break repair (DSBR) model (11). Absolute coconversion of markers between the remote *td* cleavage site and the intron insertion site (14) provides further support for the model (Fig. 1). An untested prediction of the DSBR model is that the DNA sequences encoding the catalytic core of the intron are *not* required for intron transfer, if the endonuclease is provided from a different source. We therefore wished to examine the role of intron sequences in the gene conversion event.

Previous studies showed that ORF-deleted introns were transferred successfully to In$^-$ alleles when the endonuclease was provided *in trans* (22). These results suggested that the endonucleases may be able to transfer sequences unrelated to the intron. To test whether foreign sequences within the intron core are compatible with intron inheritance and to explore the possibility of simplifying our detection methods for transfer of sequences, a 4-kb *lacZ* or a 1.2-kb *kan*^{R} fragment was cloned into the intron core of pA*td*SG2 to generate pA*td*SG2*lacZ* and pA*td*SG2*kan*^{R}, respectively. These constructs were each tested for their ability to act as intron donors when I-TevI was supplied *in trans* from compatible plasmid pK-ORF$^+$. Following

infection with T4SAΔ9*td*ΔIn phage (for pA*td*SG2*lacZ*), or induction of λ*td*ΔIn lysogens (for pA*td*SG2*kan*^{R}), marker-containing phage progeny were scored (Fig. 4). In the presence of pK-ORF$^+$, high frequency transfer of introns containing either the *lacZ* or *kan*^{R} marker occurred (69.1% and 59.3%, respectively), whereas in controls containing pK-ORF$^-$ recombination frequencies were less than 0.1% (Fig. 4A and B, constructs 1). These results indicated that, in accord with the DSBR model, the transfer mechanism is tolerant of large insertions within the intron.

The ability of foreign sequences in the intron to be transferred efficiently when I-TevI was supplied *in trans*, prompted a similar experiment to test whether marker sequences could be transferred in the absence of intron core sequences. Crosses were performed as described above, with either the *lacZ* or *kan*^{R} genes inserted directly between the exons. Thus *ca.* 66% of the progeny phage acquired *lacZ* from pA*td*Δ*InlacZ* when I-TevI was supplied *in trans* from pK-ORF$^+$, whereas < 0.1% progeny phage inherited *lacZ* in the absence of I-TevI, when pK-ORF$^-$ was used (Fig. 4A, construct 2). These results demonstrate that the intron core is not required for I-TevI-dependent sequence transfer. To verify that marker-gene transfer had indeed occurred into the *td* gene, the Td phenotype was analyzed (Fig. 4C). If *lacZ* had been inserted at a position other than *td*, the phage would remain Td$^+$. Since all Lac$^+$ phage were found to be Td$^-$, it appears that *lacZ* had inserted appropriately into the *td* gene.

To discount the possibility that *lacZ* provides information that can substitute for the catalytic intron sequences needed for mobility, the exon-*kan*^{R} construct (pA*td*Δ*Inkan*^{R}) was similarly assayed for marker inheritance. Efficient I-TevI-dependent transfer of the *kan*^{R} marker (73%) demonstrates that as for *lacZ*, intron core sequences are not required for transfer (Fig. 4B, construct 2). While supporting the DSBR model for intron mobility, the intron-independence of sequence inheritance promoted by the endonucleases has bearing on the evolution of homing elements, as discussed below.

DISCUSSION

A Distinctive Subclass of Intron-encoded Endonucleases

Although the overall similarities between the mobility phenomenon in pro- and eukaryotes imply involvement of the same basic site-specific recombination mechanism, the differences between the intron-encoded endonucleases indicate that the phage enzymes form a unique subclass. The phage endonucleases were first set apart from the eukaryotic enzymes by a relaxed specificity (ref. 22; Figs. 2C and 3C; D. B-P., unpublished cleavage data) and by the absence of the conserved dodecamer peptide motif, shared by the yeast enzymes (3, 17, 38, 39). The current study corroborates the distinctive nature of the phage endonucleases. First, the cleavage sites are displaced from the intron insertion sites in either the 5′ or 3′ direction. Cleavages are 23−26 nt 5′ (Fig. 2) and 13−15 nt 3′ (Fig. 3) to the intron insertion sites for I-TevI and I-TevII, respectively. Second, cleavage generates 2-nt (Fig. 3) or 3-nt (Fig. 2) 3′-OH extensions. In addition, preliminary experiments have shown that complementary oligonucleotides containing 24 nts surrounding the *td* intron insertion site are capable of directing cleavage into foreign DNA sequences. This suggests that the natural cleavage site is not absolutely required for binding of I-TevI, and that the binding site and cleavage site are functionally distinct. These properties of the phage endonucleases are in striking contrast to the *S. cerevisiae* (ω and aI4α) and *P. polycephalum* enzymes, which recognize sequences surrounding the insertion site and cleave at or near this site to generate 4-nt 3′ extensions (15, 20, 24, 26).

Support for the DSBR Model for Intron Mobility

Our previous demonstration of 100% coconversion of markers within 25 nt 5′ to the *td* intron provides strong support for the DSBR model for intron inheritance when viewed in the context of the I-TevI cleavages displaced 23−26 nt 5′ of the insertion site (Figs. 1 and 2). Thus, in agreement with the model, exonucleolytic degradation must proceed from the upstream cleavage to beyond the point of intron insertion to allow a successful intron transfer event. The effect of the eccentric cleavages is, therefore, to extend the transfer boundary of the intron so that residues either 5′ (*td*) or 3′ (*sun*Y) are always coinherited. Obligatory coconversion may represent a mechanism which ensures inheritance of exon sequences involved in 5′ and 3′ splice site selection (the P1 and P10 pairings, respectively—reviewed in refs. 1, 5). Such a mechanism would have important implications for intron transposition to new loci. On one hand the success of a transposition event would be limited by the compatibility of imported exon sequences with gene function; on the other hand, these exon nucleotides would ensure the splicing proficiency of the newly acquired intron.

The demonstrations that large insertions within the intron are successfully transferred, and that the intron core sequences are dispensible to mobility provide further compelling support for the applicability of the DSBR pathway to the intron transfer process (Fig. 4). Thus, the endonucleases stimulate recombination by creating ends that invade homologous donor sequences and then initiate repair synthesis using the donor strands as template. This results in the passive inheritance of the intervening non-homologous donor segment, which simply serves as the template for DNA synthesis (Fig. 1). The model therefore predicts the sequence-independence of the non-homologous stretch of donor DNA that is transferred, as substantiated in our studies. The findings that DNA sequences encoding the intron core have no active role in the recombination events that promote intron transfer support the argument that the endonucleases rather than the introns themselves are the primordial mobile elements (7, 37).

Evolution of Mobile Introns

Phylogenetic observations further support the hypothesis that the ORFs are mobile and that they have evolved independently of the intron core. The high degree of relatedness of the core structure elements of the various members of the group I intron family (2) is at variance with the distinctive nature and positioning of the ORFs (1), implying an ancestral relationship of the core structures contrasted with the independent origin of the ORFs. Indeed, even among the three T4 introns, with core structure sequences sufficiently similar that they undergo homologous recombination (40, 41), the ORFs are dissimilar with respect to both sequence and intron location (42). Strong circumstantial evidence for the independent evolution of structural and coding regions of the introns and for ORF mobility is also provided by the *Neurospora* mitochondrial intron NDI. This intron contains two different ORFs located at two different positions in closely related species (43).

Self-splicing introns can be envisioned to provide a genetically silent locus for invading ORFs (Fig. 5, step 1), with those that encode endonucleases in turn furnishing the means to propagate the composite intron. Several issues are, however, raised by the

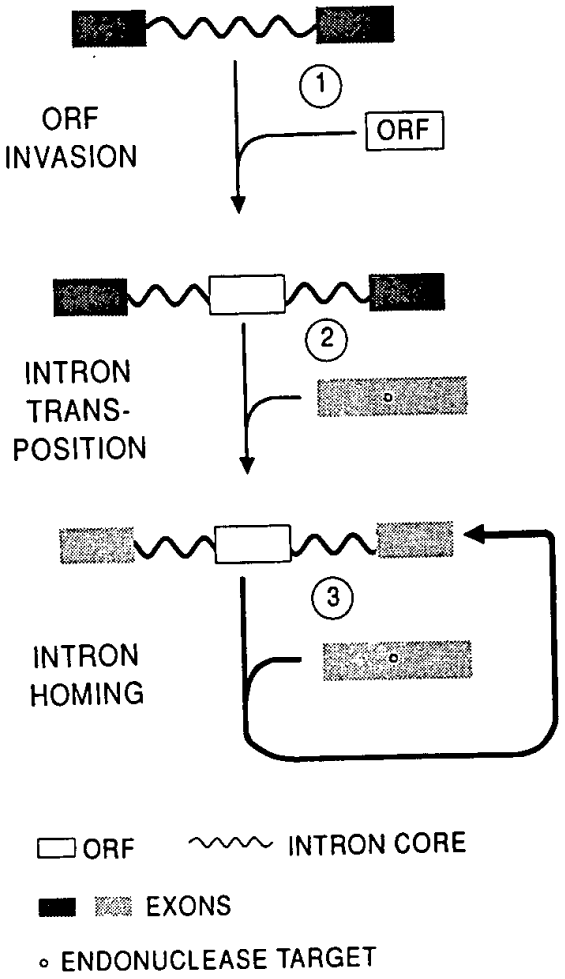


Figure 5. Model for the evolution of mobile introns. The three proposed steps are described in the text. Step 3 is depicted with a heavy arrow to emphasize the efficiency of intron homing relative to the rare events of ORF invasion (step 1) and intron transposition (step 2).

ORF invasion hypothesis. First, one must assume that successful colonization of the intron by the ORF is consistent with some degree of splicing function. Evolution toward efficient splicing may follow by adaptation of RNA structure elements or acquisition of maturase-type function by the endonucleases—the latter scenario could account for the latent maturase activity of the aI4a endonuclease (15, 20). Second, the endonuclease-containing intron must become capable of efficient homing. How might the site-specific endonuclease acquire the ability to initiate the homing process by cleavage of intronless variants? It is difficult to rationalize the endonuclease itself evolving to cleave the intronless target. As an alternative hypothesis we propose that exon target recognition might involve intron movement to a new, naturally-occurring cleavage site (Fig. 5, step 2), with transposition occurring by the same mechanism as intron homing, albeit at extremely low frequency (6). The endonuclease is then appropriately positioned to promote efficient homing by cleaving intronless alleles at its new locus (Fig. 5, step 3).

The essence of the model is that mobile (i.e. homing) introns in modern genomes are translocated descendants of introns from sites at which they acquired endonuclease ORFs. Intron transposition must, however, be a very rare event, limited by the requirements for exon homology (22), and splicing function (i.e. the appropriate intron-exon contracts; refs. 1, 5) in the new environment (6, 14, 44). Nevertheless, the occurrence of closely-related mobile introns in different genes provides circumstantial evidence consistent with occasional transposition by the endonuclease-mediated pathway. For example, the phage introns are likely to have evolved from a common ancestral intron that may have acquired different ORFs, which account for the present intron locations. As another example, an intron that is highly homologous to the ω intron in the rRNA gene of *S. cerevisiae* is found in the ATPase subunit 9 gene of the related yeast *Kluyveromyces fragilis* (45). The endonuclease is envisaged to have cleaved a foreign target sequence, stimulating a rare repair event that resulted in translocation of the composite intron (45).

If ORF invasion occurred via the same enzymatic mechanism as intron homing, one might expect that in some mobile introns vestigial similarities exist between intron sequences flanking the ORF and exon sequences around the intron insertion site. Indeed, the *sun*Y intron insertion site (IS) between the exons matches sequences immediately 5′ to the start of the intron ORF (ATG marked with dots) in 14 of 17 positions (underscored):

```
                  IS                CS
                  |                 |
Exons    TATGAGTATGAAGTGAACACGTTATTCA
         ::::: :: :: ::::: : :
Intron   TATGAATACGAGGTGAAACGATGAAATG
                             ...
```

Remarkably, in the above alignment the start codon of the ORF coincides precisely with the cleavage site (CS) of I-TevII. Despite the general difficulty of solving such evolutionary conundrums, the mobile phage introns are amenable to future tests of some of these ideas.

ACKNOWLEDGEMENTS

We thank Doris Dixon for constructing M13*sun*YΔIn and pA*td*Δ*Inkan*^{R}, Renee Schroeder for stimulating discussions, Mary Bryk and Jill Salvo for comments on the manuscript, David Shub and Nancy Kleckner for providing strains, and Bernard Dujon, Claude Lemieux and Volker Vogt for sharing results prior to publication. This work was supported by grants from the National Science Foundation (DMB8502961) and the National Institutes of Health (GM39422).

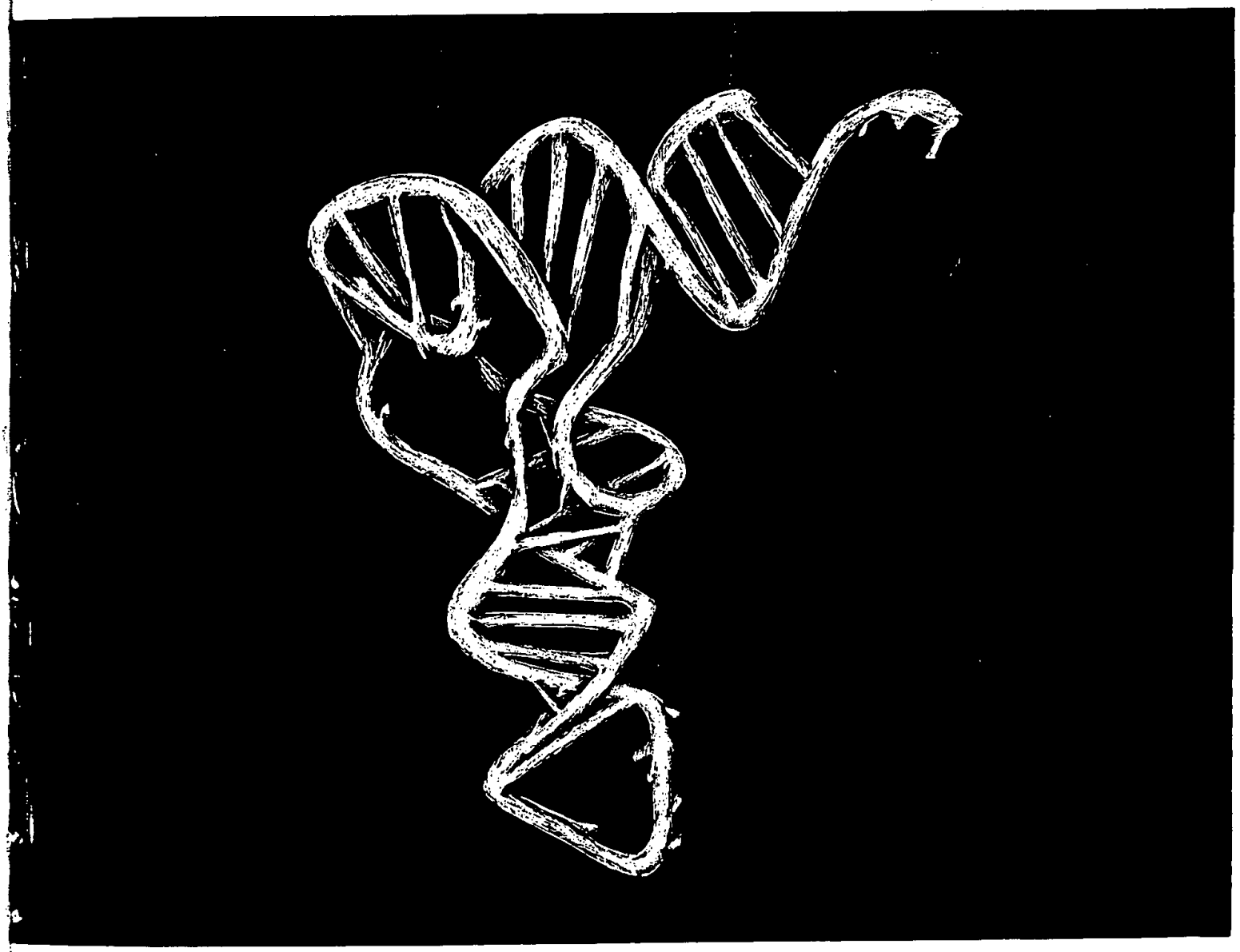
REFERENCES

1. Burke, J.M. (1988) *Gene* 73, 273–294.
2. Cech, T.R. (1988) *Gene* 73, 259–271.
3. Michel, F., Jacquier, A. and Dujon, B. (1982) *Biochimie* 64, 867–881.
4. Davies, R.W., Waring, R.B., Ray, J.A., Brown, T.A. and Scazzocchio, C. (1982) *Nature* 300, 719–724.
5. Cech, T.R. (1990) *Ann. Rev. Biochem.* 59, in press.
6. Dujon, B. (1989) *Gene* 82, 91–114.
7. Perlman, P.S. and Butow, R.A. (1989) *Science* 246, 1106–1109.
8. Dujon, B. and Jacquier, A. (1983) In Schweyen, R.J., Wolf, K. and Kaudewitz, F. (eds.), Mitochondria. Walter de Gruyter, Berlin, pp. 389–403.
9. Zinn, A.R. and Butow, R.A. (1985) *Cell* 40, 887–895.
10. Orr-Weaver, T.L., Szostak, J.W. and Rothstein, R.J. (1981) *Proc. Natl. Acad. Sci.* 78, 6354–6358.
11. Szostak, J.W., Orr-Weaver, T.L., Rothstein, R.J. and Stahl, F.W. (1983) *Cell* 33, 25–35.

12. Dujon, B., Belfort, M., Butow, R.A., Jacq, C., Lemieux, C., Perlman, P.S. and Vogt, V.M. (1989) *Gene* **82**, 115–118.
13. Jacquier, A. and Dujon, B. (1985) *Cell* **41**, 383–394.
14. Bell-Pedersen, D., Quirk, S., Aubrey, M. and Belfort, M. (1989) *Gene*, **82**, 119–126.
15. Wenzlau, J.M., Saldanha, R.J., Butow, R.A. and Perlman, P.S. (1989) *Cell* **56**, 421–430.
16. Muscarella, D.E. and Vogt, V.M. (1989) *Cell* **56**, 443–454.
17. Dujon, B. (1980) *Cell* **20**, 185–197.
18. Macreadie, I.G., Scott, R.M., Zinn, A.R. and Butow, R.A. (1985) *Cell* **41**, 395–402.
19. Colleaux, L., d'Auriol, L. Betermier, M., Cottarel, G. Jacquier, A. Galibert, F. and Dujon, B. (1986) *Cell* **44**, 521–533.
20. Delahodde, A., Goguel, V., Becam, A.M., Creusot, F., Perea, J., Banroques, J. and Jacq, C. (1989) *Cell* **56**, 431–441.
21. Lemieux, B., Turmel, M. and Lemieux, C. (1988) *Mol. Gen. Genet.* **212**, 48–55.
22. Quirk, S.M., Bell-Pedersen, D. and Belfort, M. (1989) *Cell* **56**, 455–465.
23. West, D.K., Changchien, L.-M., Maley, G.F. and Maley, F. (1989) *J. Biol. Chem.* **264**, 10343–10346.
24. Colleaux, L., d'Auriol, L., Galibert, F. and Dujon, B. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 6022–6026.
25. Monteilhet, C., Perrin, A., Thierry, A., Colleaux, L. and Dujon, B. (1990) *Nucleic Acids Res.* **18**, 1407–1413.
26. Muscarella, D.E., Ellison, E.L., Ruoff, B.M. and Vogt, V.M. *Mol. Cell Biol.*, in press.
27. Hensgens, L.A.M., Bonen, L., deHaan, M., van der Horst, G. and Grivell, L.A. (1983) *Cell* **32**, 379–389.
28. West, D.K., Belfort, M., Maley, G.F. and Maley, F. (1986) *J. Biol. Chem.* **261**, 13446–13450.
29. Shub, D.A., Xu, M.-Q., Gott, J.M., Zeeh, A. and Wilson, L.D. (1987) *Cold Spring Harbor Symp. Quant. Biol.* **52**, 193–200.
30. Chang, A.C.Y., and Cohen, S.N. (1978) *J. Bact.* **134**, 1141–1156.
31. Salvo, J.L., Coetzee, T, and Belfort, M. (1990) *J. Mol. Biol.* **211**, 537–549.
32. Simons, R.W., Houman, F. and Kleckner, N. (1987) *Gene* **53**, 85–96.
33. Kunkel, T.A., Roberts, J.D. and Zakour, R.A. (1987) *Meth. Enzymol.* **154**, 367–382.
34. Rosenberg, M., Ho, Y.-S. and Shatzman, A. (1983) *Meth. Enzymol.* **101**, 123–138.
35. Young, R.A. and Davis, R.W. (1985) In Setlow, J. and Hollaender, A. (eds.), Genetic Engineering, Vol. 7, Plenum, New York, pp. 29–41.
36. Quirk, S.M., Bell-Pedersen, D., Tomaschewski, J., RÜger, W. and Belfort, M. (1989) *Nucleic Acids Res.* **17**, 301–315.
37. Lambowitz, A.M. (1989) *Cell* **56**, 323–326.
38. Chu, F.K., Maley, G.F., West, D.K., Belfort, M. and Maley, F. (1986) *Cell* **45**, 157–166.
39. Tomaschewski, J. and Rüger, W. (1987) *Nucleic Acids Res.* **15**, 3632.
40. Hall, D.H., Liu, Y., and Shub, D.A. (1989) *Nature* **340**, 574–576.
41. Bryk, M., and Belfort, M. (1990) *Nature*, in press.
42. Shub, D.A., Gott, J.M., Xu, M.-Q., Lang, B.F., Michel, F., Tomaschewski, J., Pedersen-Lane, J. and Belfort, M. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 1151–1155.
43. Mota, E.M. and Collins, R.A. (1988) *Nature* **332**, 654–656.
44. Belfort, M. (1989) *Trends Genet* **5**, 209–216.
45. Dujon, B., Colleaux, L., Jacquier, A., Michel, F., and Monteilhet, C. (1986) In Wickner, R.B., Hinnebush, A., Lambowitz, A.M., Gunsalus, I.C. and Hollaender, A. (eds.) Extrachromosomal Elements in Lower Eukaryotes. Plenum Press, New York. pp. 5–27.
46. Chu, F.K., Maley, G.F., Maley, F. and Belfort, M. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 3049–3053.
47. Hall, D.H. (1967) *Proc. Natl. Acad. Sci. USA* **58**, 584–591.

VOLUME 18 NUMBER 13 JULY 11, 1990

Nucleic Acids Research



IRL PRESS
at
OXFORD UNIVERSITY PRESS
Oxford New York Tokyo

ISSN 0305 1048 Coden NARHAD